Real-time People Counting Method with Surveillance Cameras Implemented on Embedded System

Hsiang-Chieh Chen, Ya-Ching Chang, Nai-Jen Li, Cheng-Feng Weng, and Wen-June Wang

***Abstract*—This work presents an image-based method to estimate the number of people crowd in a large-spaced indoor environment. The method is implemented on embedded system using a digital signal processor. Moving people in each video frame are first extracted using background subtraction method. Additionally, an adaptation scheme is employed to update background model for conditions of sudden lighting changes and abandoned objects. The extracted foreground objects are further accumulated as weighted pixels to estimate the number of people, where the weights are automatically calculated by considering camera settings. Experimental results demonstrate that the proposed method performs well on estimating the number of people in the video surveillance.**

***Index Terms*—Digital signal processor, people counting, video surveillance**

I. Introduction

VISUAL surveillance systems with image processing techniques have recently developed in many applications on embedded-based platforms due to the advancement of technology in industries. Generally a surveillance system could be designed as a distributed computing structure using a server and some digital signal processors (DSPs) for varies applications. Besides, a DSP also provides the capability of complicated computation to replace a PC in specific applications. In [1], the data of images are first extracted by a PC and thus transmitted to DSPs for consequent processing. An intelligent lighting control system implemented on embedded platform with novel image processing algorithms is introduced in [2]. The position and number of occupants analyzed using a DSP are employed to control indoor lights. A face detection system based on a DSP is presented [3]. The research introduced in [4] not only performs a recognition algorithm but also proposes some schemes including data packaging and loop optimization to improve the speed of program. A moving object detection method is realized on a DSP called TMS320DM643 and its corresponding code is particularly rewritten for better performance [5].

This work proposes a novel people counting method for estimating the number of people crowd in large spaces such as lobbies, wholesale- and super-stores. Thus the estimated results could be used in different applicable scenarios, such as lighting and air conditioning control, for advancement of our lives or better efficiency of energy use. This paper is organized as follows. Section II introduces a weighting model related to various monitoring distances and an adjustable human template for counting people from extracted foreground objects. Section III describes the main procedure of our algorithm. Finally, the experimental results and conclusions are respectively presented in Sections IV and V.

II. Weighting Model And Human Template

In this work, the number of people present is mainly calculated from the accumulated pixels of foreground objects. Unfortunately, objects in each image capture probably have optical distortion. Fig.1 illustrates the condition that an identical person standing at different positions, A' and B', ahead of a camera forms different image sizes. Thus the weighting model presented below is employed to regulate those formed image sizes to identical results while the foreground pixels are accumulated by weightings.

A. Weighting Model

Fig. 2 shows a lateral view of the proposed method in which θ and ϕ^V respectively denote the angle of depression and a half of vertical view angle. The camera is set at point H. Fig. 3 illustrates the relationship between the capture frame of surveillance video and real world. The trapezoidal area ABCD in real coordinator is monitored and captured as an image EFCD. In other words, the area ABCD is a real ground region and is projected on the screen of image frame EFCD on the green line $\overline{KJ}$ in Fig. 2. First, a half of vertical view angle ϕ^V is calculated using (1).

$$\phi^V = \tan^{-1}(H_{CCD} / 2 f_{CCD}), \qquad (1)$$

where H_{CCD} is the height of CCD device in a camera and f is the focal length of used lens. Assume the captured image is sized by 320 pixel×240 pixel that is frequently used in surveillance systems and refer to Fig. 2 again. The following equations could be simply derived.

$$\overline{HJ} = \overline{HI} / \cos(90° - \theta - \phi^V), \qquad (2)$$

This work was supported by the Bureau of Energy, Taiwan, under the project of Key Technology Development of Smart Energy Network and Energy Saving Control.

Hsiang-Chieh Chen is with the Industrial Technology Research Institute, Hsinchu 310, Taiwan (phone: 886-3-5914927; fax: 886-3-5820050; email: hcchen@itri.org.tw).

Ya-Ching Chang is with the Industrial Technology Research Institute (email: gloriachang@itri.org.tw).

Nai-Jen Li is with the Department of Electrical Engineering, National Central University (email: 985401029@cc.ncu.edu.tw).

Wen-June Wang is with the Department of Electrical Engineering, National Central University (email: wjwang@ee.ncu.edu.tw).

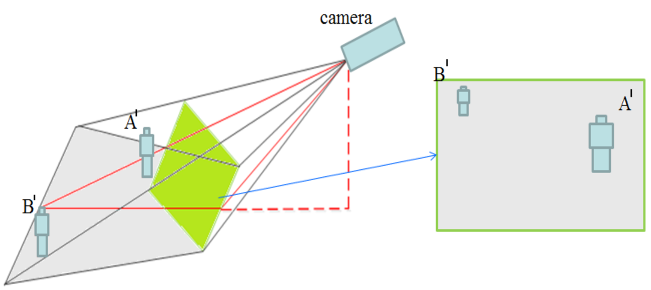


Fig. 1. People at different positions result in different sizes in an image.

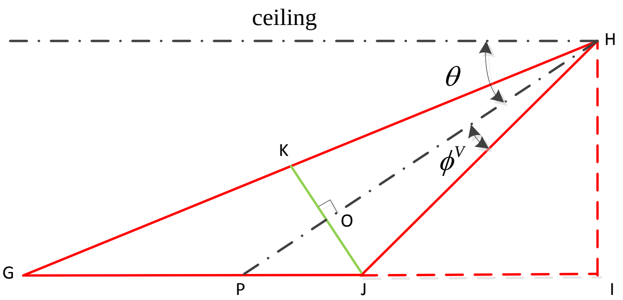


Fig. 2. The lateral view of setting up a camera.

$$\overline{OJ} = \overline{HJ}\sin(\phi^V), \tag{3}$$

$$\overline{OH} = \overline{HJ}\cos(\phi^V), \tag{4}$$

$$\overline{IJ} = \overline{HI}\tan(90^\circ - \theta - \phi^V), \tag{5}$$

and the real length $\overline{PJ}$ projecting on $\overline{OJ}$ with every row y is as follows:

$$L_y = \overline{HI}\tan\left[90^\circ - \theta - \tan^{-1}\left(\frac{\overline{OJ}(y-121)}{\overline{OH}(120)}\right)\right] - \overline{IJ} - \sum_{j=y+1}^{240} L_j, \tag{6}$$

where $y = 121,122,\ldots,240$, and L_{240} is the real length $\overline{PJ}$ in Fig. 2 projecting on $\overline{OJ}$ at $y = 240$. Fig. 4 illustrates that if there are two row positions $y = 240$ and $y = 239$ in frame, L_{239} and L_{240} which present the real length of pixels at $y = 239$ and $y = 240$ could be obtained. Similarly, the real length $\overline{PG}$ projecting on $\overline{OK}$ with every row y is as follows:

$$L_y = \overline{HI}\tan\left[90^\circ - \theta - \tan^{-1}\left(\frac{\overline{OK}(121-y)}{\overline{OH}(120)}\right)\right] - \overline{IJ} - \sum_{j=y+1}^{240} L_j, \tag{7}$$

where $y = 1,2,\ldots,120$. The following derivation is to calculate a weight value for each row based on the results of Eqs. (6) and (7). From the ground region AGJD in Fig. 3, we extend $\overline{AD}$ and $\overline{GH}$ to point Q to form a triangle AGQ as shown in Fig. 5 that illustrates the top view of Fig. 3. Here ϕ^H denotes a half of horizontal view angle. By applying trigonometry formulas, we can obtain W_y which plays a role is like L_y as follows:

$$W_y = \left(\overline{JQ} + L_y\right)\tan(\phi^H), \tag{8}$$

where

$$\phi^H = \tan^{-1}(\overline{DJ}/\overline{JQ}), \tag{9}$$

$$\overline{DJ} = \overline{EK}, \tag{10}$$

and

$$\overline{JQ} = \overline{DJ}\sum_{j=1}^{240} L_j/(\overline{AG} - \overline{DJ}). \tag{11}$$

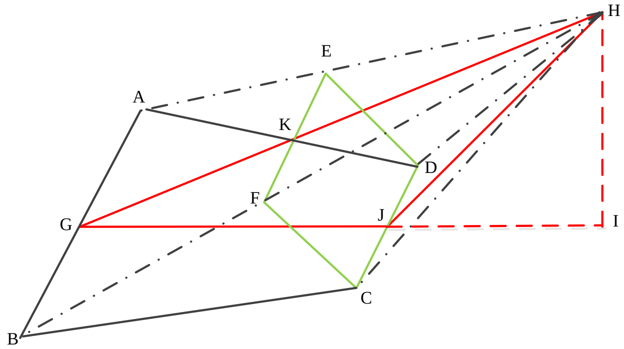


Fig. 3. The relationship between the image frame and real world.

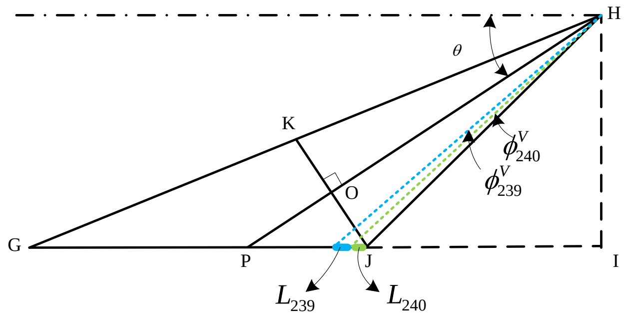


Fig. 4. The length in real world at $y = 239$ and $y = 240$.

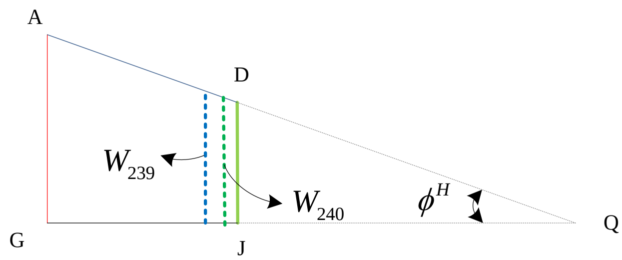


Fig. 5. Top view illustrating the real width at $y = 239$ and $y = 240$.

Finally, the weighting model is presented in the following equation:

$$\text{weight}(y) = W_y/W_{240},\ y = 1,2,\ldots,240. \tag{12}$$

B. *Human Template*

An adjustable human template that presents a virtual person standing at $y = 240$ in proposed to estimate the number of people. The human template is plotted as $\overline{JV}$ in Fig. 6, in which

$$\overline{IJ} = \overline{HI} \times \tan(90^\circ - \theta - \phi^V), \tag{13}$$

and

$$\overline{HJ} = \overline{HI} \times \cos(90^\circ - \theta - \phi^V). \tag{14}$$

By applying similar triangles between ΔHIJ and ΔJUV, the $\overline{UV}$ is derived as

$$\overline{UV} = \overline{JV} \times \sin(90^\circ - \theta - \phi^V). \tag{15}$$

Then we get $\overline{JY}$ that $\overline{JV}$ projects on $\overline{JO}$ using (16).

$$\overline{JY} = \overline{JO} - \overline{OH} \times \tan(\phi^V - \alpha), \tag{16}$$

where

$$\alpha = \tan^{-1}[\overline{UV}/(\overline{HJ} - \overline{HJ} \times \cos(90^\circ - \theta - \phi^V))], \tag{17}$$

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

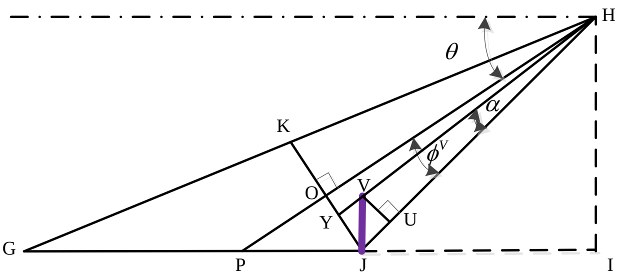


Fig. 6. The human template $\overline{JV}$ standing at y = 240.

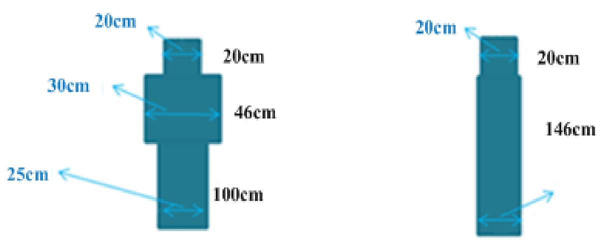


Fig. 7. Two typical human templates.

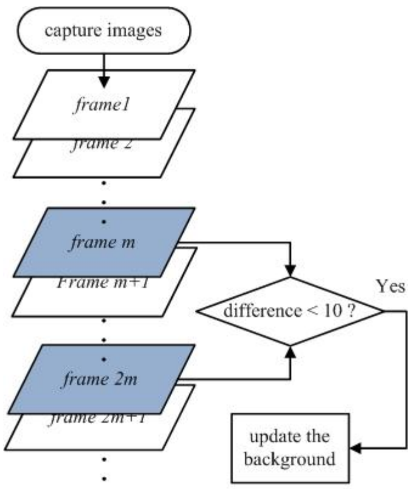


Fig. 8. Flowchart of background updating scheme.

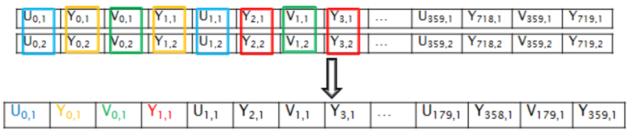


Fig. 9. Down-sizing method for YUV 4:2:2 images.

$$\overline{JO} = \overline{HJ} \times \sin(\phi^V), \tag{18}$$

and

$$\overline{OH} = \overline{HJ} \times \cos(\phi^V). \tag{19}$$

Fig. 7 shows two typical human templates used in the following experiments. Suppose the height of a person is 166 cm that is the average from statistical method. The number of pixels for 1 cm is obtained as

$$\mathrm{p} = (120\overline{JY})/(166\overline{JO}). \tag{20}$$

Therefore, a virtual person of a human template in real size is transformed in pixel unit and then the template in pixel form is weighed by the same weighting model. Finally, we accumulate all weighed pixels of adjustable human template to consider as a unit sample of human template N_{sample}. The type of the adjustable human template would be changed because of the movement direction that center of gravity of accumulated blob in foreground shifts horizontally or vertically and the template would change to be the left or right one in Fig. 7, respectively.

III. Main Algorithm

The main object of this work is to estimate the number of people using the above-mentioned weighting model and human templates. Additionally the proposed algorithm is implemented on embedded system with a DSP. The main procedure of our algorithm is described in the following four subsections.

A. *Image pre-processing*

The image pre-processing provides useful features from captured images. A background model is constructed using the first captured image for the background subtraction method [6] to extract foreground objects. Morphological processing including erosion and dilation operations is employed to eliminate foreground noises.

B. *Background Updating Scheme*

A robust background model used to extract a desired foreground is critical. In this work, a simple and widely-used method [7] is utilized to update the background model to achieve a good balance between performance and computational complexity. The method first saves the *m-th* and *2m-th* image frames from sequence captures. By subtracting these two images pixel by pixel, the corresponding pixels whose results are smaller than a predefined threshold have to be updated. Fig. 8 plots the flowchart of background updating scheme, where the predefined threshold is set as 10).

C. *Image Down-sampling for DSP*

The format of video source is YUV 4:2:2 and captured frame is sized by 640 × 480 pixels. A downsizing scheme is necessary since the image processing takes a great deal of computation of a DSP. Fig. 9 demonstrates the down-sampling method that takes an average value of four pixels. For example, we take the average of $U_{0,1}$, $U_{0,2}$, $U_{1,1}$ and $U_{1,2}$ to be the down-sized pixel $U_{0,1}$.

D. *People Number Estimation*

Foreground objects are extracted by subtraction method and then be processed by sequence of erosion and dilation operations. Next the proposed weighting model is utilized to calculate each pixel of foreground to obtain the accumulative weighted pixels using (21).

$$N_{foreground} = \sum_{y=1}^{240} N(y) \times weight(y)^2, \tag{21}$$

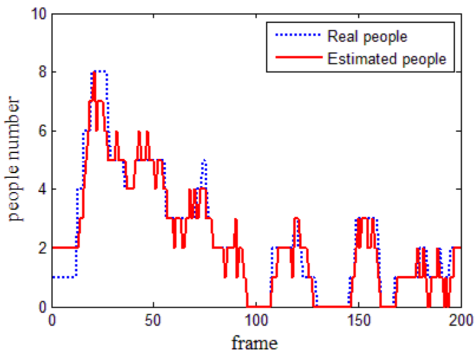
where $N_{foreground}$ is the number of the weighed pixel of foreground, $N(y)$ is the foreground pixel in the *y-th* row. Thus the number of people N_{people} is calculated

$$N_{people} = N_{foreground}/N_{sample}. \tag{22}$$

The main steps of the proposed method are summarized below.

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

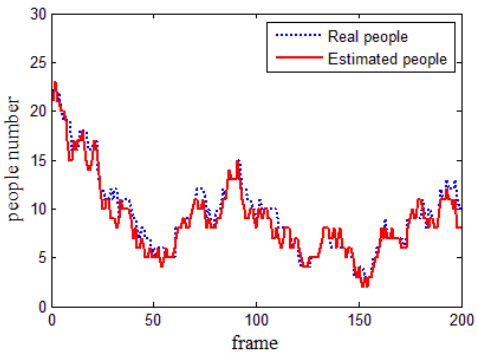
(a)



(b)

Fig. 10. The first experiment: (a) is real monitoring surroundings, and (b) is the estimation results.

(a)



(b)

Fig. 11. The second experiment: (a) is real monitoring surroundings, and (b) is the estimation results.

Step 1): Capture frame from a surveillance camera.
Step 2): Downsizing the captured frame.
Step 3): Construct background model from the first frame.
Step 4): Update background model.
Step 5): Extract foreground using background subtraction.
Step 6): Construct the weighting model and human templates.
Step 7): Accumulate the weighed pixels.
Step 8): Determine the number of people.

IV. Experimental Results

The proposed method was implemented on embedded system with a DSP called TMS320DM6437 provided by Texas Instruments. Performance of counting people was evaluated using the videos from large spaces such as lobbies and wholesale stores. Situations with different crowd densities were tested to observe the accuracy of people counting, as shown in Figs. 10(a) and 11(a). In the first experiment, the setup height of camera $\overline{HI}$ was 316 cm from floor and the angle of depression was 18.6°. In this experiment, 200 image frames were captured as the frame rate was set as 1 frame per second. Fig. 10(b) plotted the counting results, where the blue line and red one respectively denoted the real and estimated number of people. The mean absolute error (MAE) calculated from those frames was 0.45 people per frame. Similarly, Fig. 11(b) demonstrated the second experiment whose MAE was 0.92 people per frame. Either in the first experiment with fewer people or in the second experiment with more people, the counting MAE in these cases was less than 1 people per frame. The experimental results tested in this work evidently demonstrated the well performance and accuracy of people counting.

V. Conclusions

This work addresses a method for estimating the number of people. Additionally the method is further implemented on embedded system. To deal with optical deformation, a weighting model related to positions ahead of cameras is first introduced and is applied to calculate weighted foreground pixels. Consequently human templates are employed to count people from the extracted foreground objects or groups. From the experiments with different situations, the proposed method has good performance on estimating the number of occupants.

References

[1] C. Yin and D. Zhao, “The Research on Target Recognition and Image Tracking System Based on High-speed DSP,” in *Proc. IEEE Int. Conf. Comput. Sci. Educ.*, Nanning, China, Jul. 2009, pp. 721-726.

[2] X.-C. Li, L. Zhao and S.-H. Zhao, “The Establishment and Application of Based on DSP Image Processing Dynamic Multi-threshold Template,” in *Proc. IEEE Int. Conf. Futur. Comput. Commun.*, Wuhan, China, May 2010, pp. 828-832.

[3] L. Li, Y. Zhang and Q. Tian, “Multi-face Location on Embedded DSP Image Processing System,” in *Proc. IEEE Int. Conf. Image Signal Process.*, Sanya, China, May 2008, pp. 124-128.

[4] K. Li, Y. Zhu, and Y. Tian, “Implementation and Optimization of A Weed Identification Algorithm on the DSP with C64+ Core,” in *Proc. IEEE Int. Conf. Intell. Signal Process. Commun. Syst.*, Chengdu, China, Dec. 2010, pp. 1-4.

[5] R. Duan, G. Wan, D. Dong, and H. Zhou, “Design and Implementation of Embedded Moving Objects Detection System,” in *Proc. IEEE Int. Conf. Intell. Netw. Intell. Syst.*, Tianjin, China, Nov. 2009, pp. 405-408.

[6] Massimo Piccardi, “Background subtraction techniques: a review,” in *Proc. IEEE Int. Conf. Syst. Man Cyber*, Hague, Netherlands, 2004, pp. 3099-3104.

[7] N.J. Li, C.F. Chuang, Y.T. Wei, W.J. Wang, and H.C. Chen, “A video surveillance systemfor people detection and number estimation,” in *Proc. of 2012 International Conference on Fuzzy Theory and it's Applications*, Taichung, Taiwan, Nov. 2012, pp. 249-253.

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)